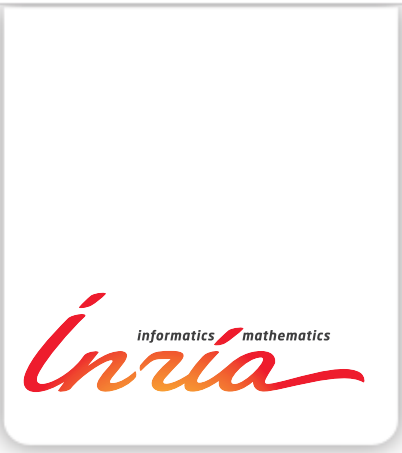
Robust Estimation of Directions-of-Arrival in Diffuse Noise Based on Matrix-Space Sparsity

Nobutaka Ito, Emmanuel Vincent, Nobutaka Ono, Shigeki Sagayama

RESEARCH REPORT
N° 8120
October 2012
Project-Teams Metiss

ISSN 0249-6399 ISRN INRIA/RR--8120--FR+ENG



Robust Estimation of Directions-of-Arrival in Diffuse Noise Based on Matrix-Space Sparsity

Nobutaka Ito*, Emmanuel Vincent, Nobutaka Ono†, Shigeki Sagayama‡

Project-Teams Metiss

Research Report n° 8120 — October 2012 — 20 pages

Abstract: We consider the estimation of the Directions-Of-Arrival (DOA) of target signals in diffuse noise. The state-of-the-art MUltiple SIgnal Classification (MUSIC) algorithm necessitates accurate identification of the signal subspace. In diffuse noise, however, it is difficult to identify it directly from the observed spatial covariance matrix. In our approach, we estimate the target spatial covariance matrix, so that we can identify the orthogonal complement of the signal subspace as its null space. We present a unified framework for modeling noise covariance in a matrix space, which generalizes four state-of-the-art diffuse noise models. We propose two alternative algorithms for estimating the target spatial covariance matrix, namely Low-rank Matrix Completion (LMC) and Trace Norm Minimization (TNM). These rely on denoising of the observed spatial covariance matrix via orthogonal projection onto the orthogonal complement of the noise matrix subspace. The missing component lying in the noise matrix subspace is then completed by exploiting the low-rankness of the target spatial covariance matrix. Large-scale experiments with real-world noise show that TNM with a certain noise model outperforms conventional MUSIC based on Generalized EigenValue Decomposition (GEVD) by 5% in terms of the precision averaged over the dataset.

Key-words: diffuse noise, DOA estimation, microphone arrays, MUSIC, matrix completion.

* Nobutaka Ito is with NTT Communication Science Labs. He performed this work as a joint PhD student with the University of Tokyo and University of Rennes 1.

† Nobutaka Ono is with the National Institute of Informatics, Japan.

‡ Shigeki Sagayama is with the Universito of Tokyo.

RESEARCH CENTRE
RENNES – BRETAGNE ATLANTIQUE

Campus universitaire de Beaulieu
35042 Rennes Cedex

Estimation Robuste de Directions d'Arrivée dans du Bruit Diffus Basée sur la Parcimonie dans un Espace Matriciel

Résumé : Nous considérons l'estimation des directions d'arrivée de sources sonores dans du bruit diffus. L'algorithme de l'état de l'art MUSIC (*MUltiple SIgnal Classification*) nécessite l'identification précise du sous-espace signal. En présence de bruit diffus, cependant, il est difficile de l'estimer directement à partir de la matrice de covariance spatiale observée. Dans notre approche, nous estimons la matrice de covariance spatiale de la source cible, de sorte à pouvoir identifier le complément orthogonal du sous-espace signal comme son espace nul. Nous présentons un cadre unifié pour la modélisation de la matrice de covariance du bruit dans un espace matriciel, qui généralise quatre modèles de bruit diffus de l'état de l'art. Nous proposons deux algorithmes pour estimer la matrice de covariance spatial de la cible, basés soit sur la complétion de matrice de rang faible soit sur la minimisation de la norme trace. Ces algorithmes reposent sur le débruitage de la matrice de covariance spatiale observée par projection orthogonale sur le complément du sous-espace matriciel correspondant au bruit. La composante manquante dans le sous-espace matriciel correspondant au bruit est alors complétée en utilisant le faible rang de la matrice de covariance spatiale de la cible. Des expériences à grande échelle montrent que, pour l'un des modèles de bruit, la minimisation de la norme trace dépasse l'approche classique par MUSIC avec décomposition en valeurs propres généralisée de 5% en terme de précision en moyenne.

Mots-clés : bruit diffus, estimation de direction d'arrivée, antenne de microphones, MUSIC, complétion de matrice.

1 Introduction

We address the DOA estimation of multiple target signals in *diffuse* noise. This has many applications epitomized by automatic camera steering [1]. For instance, we encounter diffuse noise when many people are speaking at the same time in the street or at a party. Another example is noise in a car or on a train that is caused by the vibration of the body and the windows, which constitute *surface noise sources* instead of point noise sources.

While single-source DOA estimation techniques are now established [2–6], multi-source DOA estimation has remained challenging to date. There are three main approaches, namely clustering methods [7–9], SRP-Phat [10] and its variants, and MUSIC [11, 12].

MUSIC is based on the assumption that the target signals reside in a low-dimensional *target signal subspace* of the signal space $\mathbb{C}^M$ spanned by the multichannel observed signals. It is based on null steering into the target DOAs by exploiting vectors orthogonal to the target signal subspace. Therefore, accurate identification of the target signal subspace is crucial. MUSIC can estimate DOAs even in diffuse noise in principle, because its presence does not violates the assumption that the target signals reside in a low-dimensional target signal subspace. However, the identification of the target signal subspace in diffuse noise has been difficult because of its unknown spatial correlation.

In this paper, we propose methods for estimating the target spatial covariance matrix from the observed signals, so that the orthogonal complement of the target signal subspace can be identified as its null space. We present a unified framework for modeling noise covariance in a matrix space, which generalizes four state-of-the-art diffuse noise models. We propose two algorithms for estimating the target spatial covariance matrix, namely LMC and TNM. These rely on denoising of the observed spatial covariance matrix via orthogonal projection onto the orthogonal complement of the noise matrix subspace. The missing component lying in the noise matrix subspace is then completed by exploiting the low-rankness of the target spatial covariance matrix.

We extend our preliminary paper [13] by introducing a general noise model and a new matrix completion algorithm and by performing a large-scale experiment.

The rest of this paper is structured as follows. Section 2 formulates the task considered, and reviews the state-of-the-art MUSIC. Section 3 presents the proposed noise modeling framework, and compares the fit of noise models to real-world noise. Section 4 presents the proposed algorithms for estimating the target spatial covariance matrix. Section 5 describes the experiment, and Section 6 concludes this paper.

2 Definition of DOA Estimation and Review of MUSIC

2.1 Definition of DOA estimation

We use the following notation throughout. The superscripts * and $^{\mathsf{H}}$ denote complex conjugation and Hermitian transposition, respectively. Signals are represented in the Short-Time Fourier Transform (STFT) domain as, *e.g.* $\alpha(\tau, \omega)$,

with τ and ω denoting the frame index and the angular frequency. The covariance matrix of a zero-mean vector signal $\boldsymbol{\alpha}(\tau,\omega)$ is denoted by

$$\boldsymbol{\Phi}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\tau,\omega) \triangleq \mathscr{E}[\boldsymbol{\alpha}(\tau,\omega)\boldsymbol{\alpha}^{\mathsf{H}}(\tau,\omega)], \tag{1}$$

where $\mathscr{E}[\cdot]$ is expectation.

DOA estimation is the task of estimating the DOAs of the target signals given the observed multichannel signals. When the target sources are in the far field of the array, their location is specified by two parameters, namely the azimuth and the zenith angle. We assume that the target signals are at the same height as the microphone array, and focus on the estimation of the azimuth for simplicity. However, the proposed techniques can easily be extended to the estimation of both.

Formally, we assume that an array of M microphones receives L target signals from unknown azimuths in the presence of diffuse noise. Let $\boldsymbol{s}(\tau,\omega) \in \mathbb{C}^L$ be the vector of the target signals observed at a reference point, and $\boldsymbol{x}(\tau,\omega) \in \mathbb{C}^M$ and $\boldsymbol{v}(\tau,\omega) \in \mathbb{C}^M$ be the vector of the observed signals and diffuse noise at the microphones, respectively. We denote the *steering vector* of a planewave impinging the array from a horizontal direction with azimuth ξ by

$$\boldsymbol{h}(\omega;\xi) \triangleq \begin{bmatrix} e^{-j\omega\delta_1(\xi)} & e^{-j\omega\delta_2(\xi)} & \dots & e^{-j\omega\delta_M(\xi)} \end{bmatrix}^{\mathsf{T}}, \tag{2}$$

where $\delta_m(\xi)$ is the time the planewave takes to propagate from the reference point to the m-th microphone. Therefore, denoting the target azimuths by $\Xi \triangleq \{\xi_l\}_{l=1}^L$ and defining

$$\boldsymbol{H}(\omega;\Xi) = \begin{bmatrix} \boldsymbol{h}(\omega;\xi_1) & \boldsymbol{h}(\omega;\xi_2) & \cdots & \boldsymbol{h}(\omega;\xi_L) \end{bmatrix}, \tag{3}$$

we can model the observed multichannel signal by [11]:

$$\boldsymbol{x}(\tau,\omega) = \boldsymbol{H}(\omega;\Xi)\boldsymbol{s}(\tau,\omega) + \boldsymbol{v}(\tau,\omega). \tag{4}$$

Assuming that $\boldsymbol{s}(\tau,\omega)$ and $\boldsymbol{v}(\tau,\omega)$ are mutually uncorrelated, we have the following relationship among covariance matrices:

$$\boldsymbol{\Phi}_{\boldsymbol{x}\boldsymbol{x}}(\tau,\omega) = \boldsymbol{\Phi}_{\boldsymbol{c}\boldsymbol{c}}(\tau,\omega) + \boldsymbol{\Phi}_{\boldsymbol{v}\boldsymbol{v}}(\tau,\omega), \tag{5}$$

where $\boldsymbol{\Phi}_{\boldsymbol{x}\boldsymbol{x}}(\tau,\omega)$, $\boldsymbol{\Phi}_{\boldsymbol{c}\boldsymbol{c}}(\tau,\omega) \triangleq \boldsymbol{H}(\omega;\Xi)\boldsymbol{\Phi}_{\boldsymbol{s}\boldsymbol{s}}(\tau,\omega)\boldsymbol{H}^{\mathsf{H}}(\omega;\Xi)$, and $\boldsymbol{\Phi}_{\boldsymbol{v}\boldsymbol{v}}(\tau,\omega)$ are the *observed, target, and noise covariance matrices*, respectively.

Consequently, our task is now formally defined as that of estimating Ξ given $\boldsymbol{x}(\tau,\omega)$, where the number L of target signals is assumed to be known [14].

2.2 Review of MUSIC

The observation model (4) implies that, if there are less target sources than the microphones ($L < M$), the target component $\boldsymbol{H}(\omega;\Xi)\boldsymbol{s}(\tau,\omega)$ resides in the low-dimensional target signal subspace defined by

$$\mathscr{S}(\omega) \triangleq \mathrm{span}\{\boldsymbol{h}(\omega;\xi_l)\}_{l=1}^L. \tag{6}$$

Let us denote by $\{\boldsymbol{e}_i(\omega)\}_{i=1}^{M-L}$ some basis vectors of $\mathscr{S}^{\perp}(\omega)$ (the orthogonal complement of $\mathscr{S}(\omega)$) . Each of them forms a directivity pattern with nulls at $\xi \in \Xi$:

$$|\boldsymbol{e}_i^{\mathsf{H}}(\omega)\boldsymbol{h}(\omega;\xi)|^2 = 0, \ \forall \xi \in \Xi. \tag{7}$$

Equivalently, the inverses of these directivity patterns have peaks at $\xi \in \Xi$, and so does the *narrowband MUSIC spectrum* defined by [11]:

$$f_{\mathrm{N}}(\omega;\xi) \triangleq \left[\sum_{i=1}^{M-L} |\boldsymbol{e}_i^{\mathsf{H}}(\omega)\boldsymbol{h}(\omega;\xi)|^2\right]^{-1}. \tag{8}$$

In order to integrate the information at different frequencies, the narrowband spectrum is averaged over the frequency to obtain the *wideband MUSIC spectrum*. We employ geometric averaging as in [15]:

$$f_{\mathrm{W}}(\xi) \triangleq \left[\prod_{\omega_{\min}}^{\omega_{\max}} f_{\mathrm{N}}(\omega;\xi)\right]^{\frac{1}{K}}, \tag{9}$$

Here $[\omega_{\min}, \omega_{\max}]$ denotes the frequency range of averaging, and K the corresponding number of frequency bins. The azimuth estimates $\{\hat{\xi}_l\}_{l=1}^{L}$ are obtained by picking the L largest peaks in $f_{\mathrm{W}}(\xi)$ up to the minimum angular distance of 15°.

2.3 Target signal subspace identification

It is essential in MUSIC to accurately identify $\mathscr{S}^{\perp}(\omega)$ or its basis vectors $\{\boldsymbol{e}_i(\omega)\}_{i=1}^{M-L}$ in (8). In the noiseless case, $\mathscr{S}^{\perp}(\omega)$ is obtained as the null space of $\boldsymbol{\Phi}_{\boldsymbol{xx}} = \boldsymbol{\Phi}_{\boldsymbol{cc}}$. Also, for spatially white noise, which is spatially uncorrelated and has the same power at any microphones, $\mathscr{S}^{\perp}(\omega)$ is obtained via EigenValue Decomposition (EVD) of $\boldsymbol{\Phi}_{\boldsymbol{xx}}(\tau,\omega)$ as the eigenspace of $\boldsymbol{\Phi}_{\boldsymbol{xx}}(\tau,\omega)$ corresponding to $M-L$ smallest eigenvalues. Even when noise is not spatially white, if the noise covariance matrix is known up to a scalar, $\mathscr{S}^{\perp}(\omega)$ can still be obtained via Generalized EigenValue Decomposition (GEVD) as the generalized eigenspace corresponding to the $M-L$ smallest generalized eigenvalues of the matrix pencil $(\boldsymbol{\Phi}_{\boldsymbol{xx}}(\tau,\omega), \boldsymbol{\Gamma}(\omega))$, where $\boldsymbol{\Gamma}(\omega))$ is the scaled noise spatial covariance matrix [11]. The noise covariance matrix is known *a priori* up to a scale, for some ideal noise fields such as the spherically isotropic noise field [16], which is composed of an infinite number of noise planewaves with the identical power spectrum from all three-dimensional directions in the free field. However, real-world diffuse noise can deviate from this ideal model due to the geometry of the noise sources, the room shape, and diffraction effects. Thus, the identification of $\mathscr{S}^{\perp}(\omega)$ by this approach can be unreliable.

In the following, we focus on the estimation of $\boldsymbol{\Phi}_{\boldsymbol{cc}}(\tau,\omega)$ from $\boldsymbol{\Phi}_{\boldsymbol{xx}}(\tau,\omega)$ in the presence of real-world diffuse noise, so that $\mathscr{S}^{\perp}(\omega)$ can be identified as the null space of $\boldsymbol{\Phi}_{\boldsymbol{cc}}(\tau,\omega)$.

3 Unified Noise Covariance Modeling in a Matrix Space

In order to address this estimation problem, appropriate models of $\boldsymbol{\Phi}_{\boldsymbol{cc}}$ and $\boldsymbol{\Phi}_{\boldsymbol{vv}}$ are needed. While we can exploit the low-rankness of $\boldsymbol{\Phi}_{\boldsymbol{cc}}$ assuming $L < M$, we propose a unified matrix-space model of $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega)$.

3.1 Proposed noise modeling framework

Array signal processing techniques are typically formulated in the signal space $\mathbb{C}^M$. Directional noise spans a *noise signal subspace*

$$\mathscr{N}(\omega) \triangleq \mathrm{span}\{\boldsymbol{g}_l(\omega)\}_{l=1}^{L'} \tag{10}$$

of $\mathbb{C}^M$, where $\boldsymbol{g}_l(\omega)$ are the noise steering vectors and L' the number of noise sources. Thus, if $L' < M$, noise can be eliminated by orthogonally projecting $\boldsymbol{x}(\tau,\omega)$ onto the orthogonal complement $\mathscr{N}^\perp(\omega)$ of $\mathscr{N}(\omega)$. However, diffuse noise spans whole $\mathbb{C}^M$, so that it cannot be eliminated via orthogonal projection in $\mathbb{C}^M$.

By contrast, we model $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega)$ as belonging to a *noise matrix subspace* $\mathscr{V}(\omega)$ of the vector space of Hermitian matrices

$$\mathscr{H} \triangleq \{\boldsymbol{A} \in \mathbb{C}^{M\times M} | \boldsymbol{A}^{\mathsf{H}} = \boldsymbol{A}\} \tag{11}$$

over $\mathbb{R}$. $\mathscr{H}$ is endowed with the Euclidian inner product

$$\langle \boldsymbol{A}, \boldsymbol{B}\rangle \triangleq \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^{\mathsf{H}}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) \tag{12}$$

and the Frobenius norm

$$\|\boldsymbol{A}\|_{\mathrm{F}} \triangleq \sqrt{\langle \boldsymbol{A}, \boldsymbol{A}\rangle}. \tag{13}$$

This model enables discrimination between the target signals and diffuse noise by orthogonal projection of $\boldsymbol{\Phi}_{\boldsymbol{xx}}(\tau,\omega)$ onto $\mathscr{V}^\perp(\omega)$. This is exploited in the DOA estimation algorithms proposed in Section 4.

Although $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega)$ belongs more specifically to the set of Hermitian positive semidefinite matrices, this set is not a linear space. Linear-space modeling leads to efficient algorithms by using the orthogonal projections $\mathscr{P}_\omega$ and $\mathscr{P}_\omega^\perp$ onto $\mathscr{V}(\omega)$ and its orthogonal complement $\mathscr{V}^\perp(\omega)$ as will be shown in Section 4:

$$\mathscr{P}_\omega[\boldsymbol{A}] \triangleq \sum_{i=1}^{P} \langle \boldsymbol{A}, \boldsymbol{Q}_i(\omega)\rangle \boldsymbol{Q}_i(\omega), \tag{14}$$

$$\mathscr{P}_\omega^\perp[\boldsymbol{A}] = \boldsymbol{A} - \mathscr{P}_\omega[\boldsymbol{A}], \tag{15}$$

where $\{\boldsymbol{Q}_i(\omega)\}_{i=1}^P$ denotes an orthonormal basis of $\mathscr{V}(\omega)$ with $P \triangleq \dim\mathscr{V}(\omega)$. Explicit forms of these projections are given in Section 3.2.

3.2 New interpretation of state-of-the-art diffuse noise models

The proposed matrix space model includes four state-of-the-art noise models.

3.2.1 Spatially uncorrelated noise model

Zelinski [17] proposed a method for diffuse noise suppression based on the assumption that diffuse noise is spatially uncorrelated. The assumption implies that $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega)$ is diagonal, so that the corresponding noise matrix subspace is the following M-dimensional matrix subspace

$$\mathscr{V}(\omega) = \{\boldsymbol{A} \in \mathscr{H} | \boldsymbol{A} : \text{ diagonal}\}. \tag{16}$$

$\mathscr{P}_\omega$ and $\mathscr{P}_\omega^\perp$ are given by

$$\mathscr{P}_\omega[\boldsymbol{A}] = \mathscr{D}(\boldsymbol{A}), \tag{17}$$

$$\mathscr{P}_\omega^\perp[\boldsymbol{A}] = \mathscr{O}(\boldsymbol{A}), \tag{18}$$

where $\mathscr{D}(\cdot)$ is the operation of replacing the off-diagonal entries by zeros, and $\mathscr{O}(\cdot)$ that of replacing the diagonal entries by zeros. While diffuse noise can be reasonably modeled as uncorrelated when the distances between microphones are large enough compared to the wavelength, it is highly correlated for small arrays or at low frequencies [18].

3.2.2 Fixed noise coherence model

To take noise correlation into account, McCowan *et al.* [18] assumed in their method for diffuse noise suppression that $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega)$ is fixed up to an unknown scale factor. This corresponds to the one-dimensional noise matrix subspace

$$\mathscr{V}(\omega) \triangleq \{k\boldsymbol{\Gamma}(\omega) | k \in \mathbb{R}\}, \tag{19}$$

where $\boldsymbol{\Gamma}(\omega)$ denotes the so-called *noise coherence matrix*. The orthogonal projection operators are given by

$$\mathscr{P}_\omega[\boldsymbol{A}] = \frac{(\boldsymbol{A}, \boldsymbol{\Gamma}(\omega))}{\|\boldsymbol{\Gamma}(\omega)\|_{\mathrm{F}}^2}\boldsymbol{\Gamma}(\omega). \mathscr{P}_\omega^\perp[\boldsymbol{A}] = \boldsymbol{A} - \mathscr{P}_\omega[\boldsymbol{A}]. \tag{20}$$

While the noise coherence matrix for some ideal noise fields is known [16, 18], real-world noise deviates from this ideal model as pointed out in Section 2.3.

3.2.3 Blind Noise Decorrelation (BND) model

Instead of assuming specific noise coherences, Shimizu *et al.* considered *isotropic* noise satisfying the following assumptions [19, 20]:

- equal power spectrogram $\phi_{v_m v_m}(\tau,\omega)$ at all microphones m,
- equal cross-spectrogram $\phi_{v_m v_n}(\tau,\omega)$ for microphone pairs (m,n) spaced by the same distance.

Under these assumptions, the unknown spatial covariance matrix of isotropic noise is diagonalized by a known unitary matrix for certain classes of symmetrical arrays or *crystal arrays* [19,20]. This BND technique has been applied to the estimation of the target power spectrogram in diffuse noise environments [20] and to diffuse noise suppression in the signal domain [19].

BND implies that $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega)$ is expressed as $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega) = \boldsymbol{P}\boldsymbol{\Lambda}(\tau,\omega)\boldsymbol{P}^{\mathsf{H}}$ for some unknown diagonal matrix $\boldsymbol{\Lambda}(\tau,\omega) \in \mathbb{R}^{M\times M}$ and some known unitary matrix $\boldsymbol{P} \in \mathbb{C}^{M\times M}$. (**??**) can be rewritten as $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega) = \sum_{m=1}^{M} \lambda_m(\tau,\omega)\boldsymbol{p}_m\boldsymbol{p}_m^{\mathsf{H}}$ with $\lambda_m(\tau,\omega)$ denoting m-th diagonal entry of $\boldsymbol{\Lambda}(\tau,\omega)$ and $\boldsymbol{p}_m$ the m-th column of $\boldsymbol{P}$. This implies that $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega)$ belongs to the M-dimensional subspace

$$\mathscr{V}(\omega) = \mathrm{span}\{\boldsymbol{p}_m\boldsymbol{p}_m^{\mathsf{H}}\}_{m=1}^{M}. \tag{21}$$

The projectors $\mathscr{P}_\omega$ and $\mathscr{P}_\omega^\perp$ are then given by

$$\mathscr{P}_\omega[\boldsymbol{A}] = \boldsymbol{P}\mathscr{D}(\boldsymbol{P}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{P})\boldsymbol{P}^{\mathsf{H}}, \tag{22}$$

$$\mathscr{P}_\omega^\perp[\boldsymbol{A}] = \boldsymbol{P}\mathscr{O}(\boldsymbol{P}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{P})\boldsymbol{P}^{\mathsf{H}}. \tag{23}$$

While BND is valid for a wide range of noise fields, the array geometry is restricted to certain classes.

3.2.4 Real-valued noise covariance model

We introduced a flexible model of diffuse noise applicable to arbitrary array geometries in our diffuse noise suppression method [21]. The isotropy assumptions of BND imply that $\phi_{v_m v_n}(\tau, \omega) = \phi_{v_n v_m}(\tau, \omega)$, and, by definition of the cross-spectrum, we have $\phi_{v_n v_m}(\tau, \omega) = \phi^*_{v_m v_n}(\tau, \omega)$. Therefore, $\phi_{v_m v_n} \in \mathbb{R}$, and $\boldsymbol{\Phi}_{\boldsymbol{vv}}$ belongs to the $M(M+1)/2$-dimensional noise matrix subspace

$$\mathscr{V}(\omega) \triangleq \{\boldsymbol{A} \in \mathbb{R}^{M \times M} | \boldsymbol{A}^{\mathsf{T}} = \boldsymbol{A}\}. \tag{24}$$

The projectors are given by

$$\mathscr{P}_{\omega}[\boldsymbol{A}] = \Re[\boldsymbol{A}], \tag{25}$$

$$\mathscr{P}^{\perp}_{\omega}[\boldsymbol{A}] = j\Im[\boldsymbol{A}], \tag{26}$$

where $\Re[\cdot]$ and $\Im[\cdot]$ denotes the operations of taking the real part and the imaginary part, respectively.

This model is more flexible than the spatially uncorrelated noise model and the fixed noise coherence model for spherically or cylindrically isotropic noise. Indeed, these models are real-valued [16, 22], and thus noise matrix subspaces of the real-valued noise covariance model. However, it has many parameters, which can cause overfitting.

3.3 Assessment of noise models with real-world noise

Before evaluating these four noise models for DOA estimation in Section 5, we assess their potential independently of the application as follows. We investigate two different aspects which are important to predict the performance of a certain noise model:

- the number of parameters of the model $\dim\mathscr{V}(\omega)$ compared to the number of observations $\dim\mathscr{H}$,
- the fit to real-world noise spatial covariance matrices.

Ideally, for *e.g.* twice as many parameters, we expect the fit to increase a lot. If the fit is only marginally better, the increased number of parameters is likely to result in a poorer performance due to overfitting. These two pieces of information together hence enable to predict the outcomes of subsequent experiments to a certain degree.

A 1 minute-long real-world noise signal was taken from each noise environment in the dataset described in Section 5. Note that a square array geometry was used, to which the BND model is applicable. The empirical noise covariance matrix $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\omega)$ was computed by temporal averaging of $\boldsymbol{v}(\tau, \omega)\boldsymbol{v}^{\mathsf{H}}(\tau, \omega)$ over the whole duration of each signal. We define the *discrepancy index* between the noise data and $\mathscr{V}(\omega)$ as the average over frequency and noise environments of

$$\frac{\|\mathscr{P}^{\perp}_{\omega}[\boldsymbol{\Phi}_{\boldsymbol{vv}}](\omega)\|_{\mathrm{F}}}{\|\boldsymbol{\Phi}_{\boldsymbol{vv}}(\omega)\|_{\mathrm{F}}}. \tag{27}$$

Table 1: The dimensions of $\mathscr{H}$ and each noise matrix subspace $\mathscr{V}(\omega)$ as a function of the number of microphones.

# microphones	$\mathscr{H}$	uncor	coh	BND	real
M	M^2	M	1	M	$M(M+1)/2$

This quantity is the distance between $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\omega)$ and $\mathscr{V}(\omega)$ normalized by $\|\boldsymbol{\Phi}_{\boldsymbol{vv}}(\omega)\|_{\mathrm{F}}$, where the normalization is aimed to remove the dependency on the scale of $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\omega)$.

The dimensions of $\mathscr{V}(\omega)$ and $\mathscr{H}$ as functions of M are shown in Table 1. The fixed noise coherence model (denoted by coh) has one dimension independently of M. The dimensions of spatially uncorrelated noise model (denoted by uncor) and the BND model grow linearly w.r.t. M, whereas that of the real-valued noise covariance model (denoted by real) grow quadratically.

Fig. 1 plots the discrepancy index of each model versus its dimensionality. A model closer to the origin is a good model that is able to fit real-world noise better with a smaller number of parameters. The real-valued noise covariance model gave the smallest discrepancy index of 0.16, but its largest dimension can lead to overfitting. In comparison, the BND model has only 0.4 time as high a dimension with an slight increase of 0.06 in the discrepancy index. Furthermore, its discrepancy index of 0.22 was lower than those of the remaining two models. Therefore, this model is expected to work quite well, provided that a crystal array geometry is available. The spatially uncorrelated noise model gave the highest discrepancy index of 0.57. This poor fit is due to the high spatial correlation of real-world noise. The fixed noise coherence model has only dimension 1, but nevertheless it fitted noise reasonably well with a discrepancy index of 0.36.

4 Target Covariance Estimation Based on Matrix-Space Sparsity

As pointed out in Section 2, it is essential in MUSIC to accurately estimate the orthogonal complement of the target signal subspace $\mathscr{S}^{\perp}(\omega)$, and this boils down to the estimation of the target spatial covariance matrix $\boldsymbol{\Phi}_{\boldsymbol{cc}}(\tau,\omega)$. In this section, we propose methods for estimating $\boldsymbol{\Phi}_{\boldsymbol{cc}}(\tau,\omega)$ from the observed spatial covariance matrix $\boldsymbol{\Phi}_{\boldsymbol{xx}}(\tau,\omega)$. As a direct consequence of the proposed unified noise model, the component of $\boldsymbol{\Phi}_{\boldsymbol{cc}}(\tau,\omega)$ lying in the orthogonal complement $\mathscr{V}^{\perp}(\omega)$ of the noise matrix subspace $\mathscr{V}(\omega)$ is easily obtained as

$$\mathscr{P}_{\omega}^{\perp}[\boldsymbol{\Phi}_{\boldsymbol{cc}}](\tau,\omega) = \mathscr{P}_{\omega}^{\perp}[\boldsymbol{\Phi}_{\boldsymbol{xx}}](\tau,\omega), \tag{28}$$

Therefore, the problem becomes that of estimating the remaining component $\mathscr{P}_{\omega}[\boldsymbol{\Phi}_{\boldsymbol{cc}}(\tau,\omega)]$ in $\mathscr{V}(\omega)$.

In the machine learning literature, matrix completion techniques [23–26] have been proposed to recover a low-rank matrix from part of its entries. We extend these techniques to recovery of the low-rank matrix $\boldsymbol{\Phi}_{\boldsymbol{cc}}(\tau,\omega)$ from its projection $\mathscr{P}_{\omega}^{\perp}[\boldsymbol{\Phi}_{\boldsymbol{cc}}](\tau,\omega)$ under a positive semidefiniteness constraint. Here,

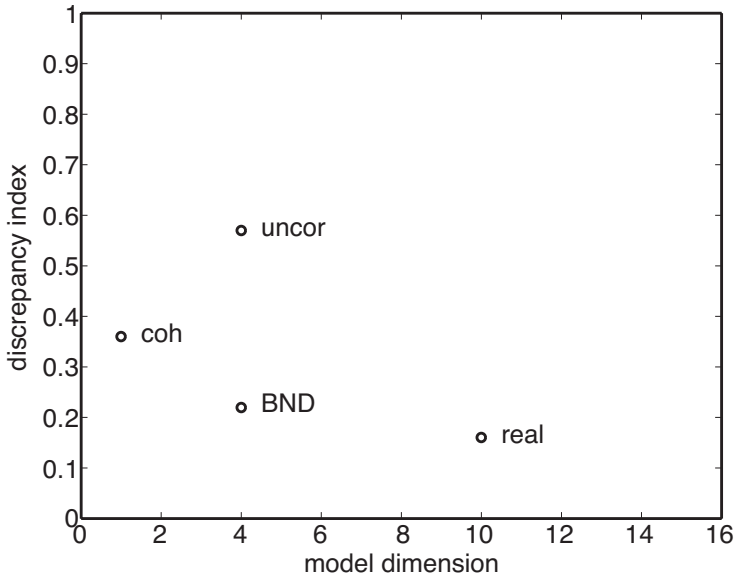


Figure 1: Discrepancy index vs. model dimension for each noise model on the noise dataset of Section 5 ($M = 4$).

the positive semidefiniteness constraint is important, so that we can identify $\mathscr{S}^{\perp}(\omega)$ as the null space of the estimated matrix.

We present two methods based on LMC (Section 4.1) and TNM (Section 4.2), which are applied in each time-frequency bin. The former is based on the main assumption that an upper bound R of $\mathrm{rank}\big(\boldsymbol{\Phi}_{\boldsymbol{cc}}(\tau,\omega)\big)$ is given, while the latter does not require that information. In the rest of this section, the time-frequency indices (τ,ω) are omitted for simplicity.

4.1 Target covariance estimation based on Low-rank Matrix Completion (LMC)

Instead of regarding $\mathscr{P}^{\perp}[\boldsymbol{\Phi}_{\boldsymbol{xx}}]$ as exactly noise-free, we leave some room for possible errors due to the misestimation of $\boldsymbol{\Phi}_{\boldsymbol{xx}}$ or to the possible inaccuracy of the noise model $\mathscr{V}(\omega)$. Specifically, we consider the following constrained minimization problem:

$$\begin{aligned} &\min_{\boldsymbol{\Phi}_{\boldsymbol{cc}}} \ \Psi_{\mathrm{comp}}(\boldsymbol{\Phi}_{\boldsymbol{cc}}) \triangleq \|\mathscr{P}^{\perp}[\boldsymbol{\Phi}_{\boldsymbol{cc}}] - \mathscr{P}^{\perp}[\boldsymbol{\Phi}_{\boldsymbol{xx}}]\|_{\mathrm{F}}^2 \\ &\text{s.t. } \boldsymbol{\Phi}_{\boldsymbol{cc}} \in \mathscr{H} \text{ positive semidefinite}, \mathrm{rank}(\boldsymbol{\Phi}_{\boldsymbol{cc}}) \leq R. \end{aligned} \tag{29}$$

We rely on the following lemma, which can be proved in line with [27]:

Lemma 1 *Consider the following optimization problem*

$$\begin{aligned} &\min_{\boldsymbol{X}} \|\boldsymbol{X} - \boldsymbol{Y}\|_F^2, \\ &\textit{s.t. } \boldsymbol{X} \in \mathscr{H} \textit{ positive semidefinite}, \textit{rank}(\boldsymbol{X}) \leq R, \end{aligned} \tag{30}$$

where $\boldsymbol{Y} \in \mathscr{H}$*. Denote the eigenvalue decomposition of* $\boldsymbol{Y}$ *by*

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{H}, \tag{31}$$

where $\boldsymbol{U} \in \mathbb{C}^{M\times M}$ *is unitary and* $\boldsymbol{\Sigma} \in \mathbb{R}^{M\times M}$ *is diagonal and composed of* $\sigma_1 \geq \cdots \geq \sigma_M$*. Then the optimal solution to (30) is given by*

$$\boldsymbol{X} = \boldsymbol{U}\max\{\boldsymbol{\Sigma}_R, 0\}\boldsymbol{U}^{H}, \tag{32}$$

where

$$\boldsymbol{\Sigma}_R \triangleq diag(\sigma_1, \ldots, \sigma_R, 0, \ldots, 0), \tag{33}$$

with $diag(\alpha_1, \ldots, \alpha_M)$ *denoting the* $M \times M$ *diagonal matrix composed of* $\alpha_1, \ldots, \alpha_M$ *and* $\max\{\cdot, 0\}$ *the operation of replacing the negative entries of a matrix with zeros.*

Compared to (30), (29) includes an orthogonal projection $\mathscr{P}_{\omega}^{\perp}$. Therefore, from the principle of Majorization-Minimization (MM) [28], we design the auxiliary function

$$\Psi_{\text{comp}}^{+}(\boldsymbol{\Phi}_{\boldsymbol{cc}}, \boldsymbol{Z}) \triangleq \Psi_{\text{comp}}(\boldsymbol{\Phi}_{\boldsymbol{cc}}) + \|\mathscr{P}[\boldsymbol{\Phi}_{\boldsymbol{cc}}] - \boldsymbol{Z}\|_{\text{F}}^{2}, \tag{34}$$

with an auxiliary variable $\boldsymbol{Z} \in \mathscr{V}$, so that

$$\Psi_{\text{comp}}(\boldsymbol{\Phi}_{\boldsymbol{cc}}) = \arg\min_{\boldsymbol{Z}} \Psi_{\text{comp}}^{+}(\boldsymbol{\Phi}_{\boldsymbol{cc}}, \boldsymbol{Z}). \tag{35}$$

Indeed, (34) can be rewritten as

$$\Psi_{\text{comp}}^{+}(\boldsymbol{\Phi}_{\boldsymbol{cc}}, \boldsymbol{Z}) = \|\boldsymbol{\Phi}_{\boldsymbol{cc}} - \boldsymbol{Y}\|_{\text{F}}^{2}, \tag{36}$$

which has the same form as (30), where

$$\boldsymbol{Y} \triangleq \boldsymbol{Z} + \mathscr{P}^{\perp}[\boldsymbol{\Phi}_{\boldsymbol{xx}}]. \tag{37}$$

The MM algorithm amounts to iteratively applying the following update rules:

$$\boldsymbol{Z} \leftarrow \arg\min_{\boldsymbol{Z}} \Psi_{\text{comp}}^{+}(\boldsymbol{\Phi}_{\boldsymbol{cc}}, \boldsymbol{Z}) \text{ s.t. } \boldsymbol{Z} \in \mathscr{V}, \tag{38}$$

$$\boldsymbol{\Phi}_{\boldsymbol{cc}} \leftarrow \arg\min_{\boldsymbol{\Phi}_{\text{cc}}} \Psi_{\text{comp}}^{+}(\boldsymbol{\Phi}_{\boldsymbol{cc}}, \boldsymbol{Z}) \tag{39}$$

$$\text{s.t. } \boldsymbol{\Phi}_{\boldsymbol{cc}}\text{: Hermitian positive semidefinite}, \operatorname{rank}(\boldsymbol{\Phi}_{\boldsymbol{cc}}) \leq R.$$

The solution to (39) is given by

$$\boldsymbol{\Phi}_{\boldsymbol{cc}} \leftarrow \boldsymbol{U}\max\{\boldsymbol{\Sigma}_R, 0\}\boldsymbol{U}^{\mathsf{H}}, \tag{40}$$

where $\boldsymbol{U}$ and $\boldsymbol{\Sigma}$ are defined in the same manner as in Lemma 1 using the eigenvalue decomposition of $\boldsymbol{Y}$. On the other hand, (38) amounts to

$$\boldsymbol{Z} \leftarrow \mathscr{P}[\boldsymbol{\Phi}_{\boldsymbol{cc}}], \tag{41}$$

because only the second term in (34) depends on $\boldsymbol{Z}$.

We iterate the algorithm until a preset maximum number of iterations K is reached or

$$\frac{\|\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k+1)} - \boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)}\|_{\mathrm{F}}}{\|\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)}\|_{\mathrm{F}}} < \epsilon, \tag{42}$$

where ϵ is a preset small constant, which means $\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)}$ has become almost constant. As for the initial value $\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(0)}$, we propose to simply use $\boldsymbol{\Phi}_{\boldsymbol{xx}}$.

The algorithm is summarized in the following:

Algorithm 1

Set $\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(0)} = \boldsymbol{\Phi}_{\boldsymbol{xx}}$, $K \geq 1$, *and* $1 \leq R < M$.
Set $k \leftarrow 0$.
repeat
$\boldsymbol{Y}^{(k)} = \mathscr{P}[\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)}] + \mathscr{P}^{\perp}[\boldsymbol{\Phi}_{\boldsymbol{xx}}]$.
Calculate the eigenvalue decomposition of $\boldsymbol{Y}^{(k)}$*:* $\boldsymbol{Y}^{(k)} = \boldsymbol{U}^{(k)}\boldsymbol{\Sigma}^{(k)}\boldsymbol{U}^{(k)H}$, *where* $\boldsymbol{U}^{(k)}$ *is unitary and* $\boldsymbol{\Sigma}^{(k)}$ *is real-valued and diagonal with the diagonal entries* $\sigma_1^{(k)}, \ldots, \sigma_M^{(k)}$ *arranged in nonincreasing order.*
$\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k+1)} = \boldsymbol{U}^{(k)} \max\{\boldsymbol{\Sigma}_R^{(k)}, 0\}\boldsymbol{U}^{(k)H}$.
$k \leftarrow k + 1$.
until $k = K$ ***or*** *(42).*

As a general property of the MM algorithm, we have the following theorem:

Theorem 1 *The sequence* $\Psi_{comp}(\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)}), k = 1, 2, \ldots$ *generated by Algorithm 1 is nonincreasing.*

Indeed, defining $\boldsymbol{Z}^{(k)}$ by

$$\boldsymbol{Z}^{(k)} \triangleq \arg\min_{\boldsymbol{Z}} \Psi^{+}(\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)}, \boldsymbol{Z}), \tag{43}$$

we have

$$\Psi(\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)}) = \Psi^{+}(\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)}, \boldsymbol{Z}^{(k)}). \tag{44}$$

Furthermore, defining $\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k+1)}$ by

$$\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k+1)} \triangleq \arg\min_{\boldsymbol{\Phi}_{\boldsymbol{cc}}} \Psi^{+}(\boldsymbol{\Phi}_{\boldsymbol{cc}}, \boldsymbol{Z}^{(k)}), \tag{45}$$

we have

$$\Psi^{+}(\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k+1)}, \boldsymbol{Z}^{(k)}) \leq \Psi^{+}(\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)}, \boldsymbol{Z}^{(k)}). \tag{46}$$

Therefore, from (44) and (46) and from

$$\Psi(\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k+1)}) \leq \Psi^{+}(\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k+1)}, \boldsymbol{Z}^{(k)}), \tag{47}$$

we have

$$\Psi(\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k+1)}) \leq \Psi(\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)}). \tag{48}$$

From Theorem 1, the convergence of $\Psi_{\mathrm{comp}}(\boldsymbol{\Phi}_{\boldsymbol{cc}})$ to a local minimum is guaranteed, because $\Psi_{\mathrm{comp}}(\boldsymbol{\Phi}_{\boldsymbol{cc}}) > 0$.

This algorithm can be regarded as an extension of Srebro's algorithm [23]. The extension is twofold. First, we consider the completion of a missing noise matrix subspace instead of missing entries. Second, we consider the completion of a complex-valued matrix with an Hermitian positive semidefiniteness constraint instead of the completion of a real-valued matrix without such a constraint.

4.2 Target covariance estimation based on Trace Norm Minimization (TNM)

We propose an alternative algorithm that does not require the upper bound R on the rank. This is advantageous because the the upper bound is not always given in practice due to an unknown number of sources and/or to reverberation.

We utilize the *trace norm* $\|\boldsymbol{\Phi}_{\boldsymbol{cc}}\|_*$, *i.e.* the sum of the singular values, to construct a cost function that favors a low-rank solution. The trace norm is known to be a convex relaxation of the rank function [24], and can be regarded as the matrix version of the popular l_1-norm for vector. Specifically, we consider the following optimization problem:

$$\begin{aligned} &\min_{\boldsymbol{\Phi}_{\boldsymbol{cc}}} \ \Psi_{\mathrm{trace}}(\boldsymbol{\Phi}_{\boldsymbol{cc}}) \triangleq \frac{1}{2}\|\mathscr{P}^{\perp}[\boldsymbol{\Phi}_{\boldsymbol{cc}}] - \mathscr{P}^{\perp}[\boldsymbol{\Phi}_{\boldsymbol{xx}}]\|_{\mathrm{F}}^2 + \mu\|\boldsymbol{\Phi}_{\boldsymbol{cc}}\|_*, \\ &\text{s.t. } \boldsymbol{\Phi}_{\boldsymbol{cc}} \in \mathscr{H} \text{ positive semidefinite,} \end{aligned} \tag{49}$$

where μ is a positive weight.

(49) can be solved efficiently by generalizing Toh's algorithm [24] to the completion of a subspace of a complex-valued matrix subject to a Hermitian positive semidefiniteness constraint. The value of μ is decreased at each iteration as proposed in [24]:

$$\mu^{(k)} = \max\{0.7^k, 10^{-4}\} \times \|\mathscr{P}^{\perp}[\boldsymbol{\Phi}_{\boldsymbol{xx}}]\|_{\mathrm{F}}. \tag{50}$$

The stopping condition is defined in the same manner as in Algorithm 1.

Algorithm 2

Set $\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(0)} = \boldsymbol{\Phi}_{\boldsymbol{cc}}^{(-1)} = \boldsymbol{\Phi}_{\boldsymbol{xx}}$; $t^{(-1)} = t^{(0)} = 1$.
$k \leftarrow 0$.
repeat
$\mu \leftarrow \max\{0.7^k, 10^{-4}\} \times \|\mathscr{P}^{\perp}[\boldsymbol{\Phi}_{\boldsymbol{xx}}]\|_F$.
$\boldsymbol{Z}^{(k)} = \boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)} + \dfrac{t^{(k-1)} - 1}{t^{(k)}}(\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)} - \boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k-1)})$.
$\boldsymbol{Y}^{(k)} = \mathscr{P}[\boldsymbol{Z}^{(k)}] + \mathscr{P}^{\perp}[\boldsymbol{\Phi}_{\boldsymbol{xx}}]$.
Calculate the eigenvalue decomposition of $\boldsymbol{Y}^{(k)}$*:* $\boldsymbol{Y}^{(k)} = \boldsymbol{U}^{(k)}\boldsymbol{\Sigma}^{(k)}\boldsymbol{U}^{(k)H}$, *where* $\boldsymbol{U}^{(k)}$ *is unitary and* $\boldsymbol{\Sigma}^{(k)}$ *is real-valued and diagonal.*
$\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k+1)} = \boldsymbol{U}^{(k)} \max\{\boldsymbol{\Sigma}^{(k)} - \mu\boldsymbol{I}, 0\}\boldsymbol{U}^{(k)H}$.
$t^{(k+1)} = \dfrac{1 + \sqrt{1 + 4t^{(k)2}}}{2}$.
$k \leftarrow k + 1$.
until $k = K$ ***or*** *(42)*

The following theorem guarantees the convergence of $\Psi_{\mathrm{trace}}(\boldsymbol{\Phi}_{\boldsymbol{cc}})$ in Algorithm 2 to a global minimum of (49).

Theorem 2 *Let* $\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)}$ $(k = 1, 2, \dots)$ *be the sequence generated by Algorithm 2. Then,*

$$|\Psi_{trace}(\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(k)}) - \Psi_{trace}(\boldsymbol{\Phi}_{\boldsymbol{cc}}^{opt})| \leq \frac{2\|\boldsymbol{\Phi}_{\boldsymbol{cc}}^{(0)} - \boldsymbol{\Phi}_{\boldsymbol{cc}}^{opt}\|_F^2}{(k+1)^2}, \tag{51}$$

where $\boldsymbol{\Phi}_{\boldsymbol{cc}}^{opt}$ *is a solution to (49).*

This can be proven in line with [24], because Lemmas 2.1–2.3 in [29] and Theorem 3 in [30] for real-valued vectors can be extended to complex-valued matrices.

4.3 Signal subspace identification

The target signal subspace $\mathscr{S}(\omega)$ and its orthogonal complement $\mathscr{S}^{\perp}(\omega)$ are calculated as follows

$$\mathscr{S} = \mathrm{span}\{\boldsymbol{u}_m\}_{m=1}^{L}, \tag{52}$$

$$\mathscr{S}^{\perp} = \mathrm{span}\{\boldsymbol{u}_m\}_{m=L+1}^{M}, \tag{53}$$

where $\boldsymbol{u}_m$ denotes the m-th column of $\boldsymbol{U}$.

5 Large-scale evaluation with real-world noise

5.1 Created dataset

We created a dataset of multichannel reverberant speech mixtures with real-world noise to evaluate the proposed algorithms. Noise was recorded in a station square in Japan with a 4-channel square array with a diameter of 5 cm [21]. To better control experimental conditions, the target components were simulated by the image method [31] implemented in Roomsimove[1]) and added to noise. In the simulation, we assumed the room dimensions to be $3.3 \times 7.8 \times 2.4$ m, the array to be at the room center, and the target sources to be at 1 m from the array and at the height of 1.2 m. The dry speech sources were taken from the ATR Japanese dataset [32]. The mixtures were 10 s long sampled at 16 kHz. The dataset includes $(3 \times 3 - 1) \times 3 \times 3 = 72$ mixtures with various values of the following four parameters:

- the number L of target sources: 2, 4, or 6,
- the azimuth separation between successive sources: 30°, 60°, or 90° (90° was not considered for $L = 6$.),
- the absorption coefficient of the walls: 0.4, 0.7, or 1.0 (*i.e.* reverberation time RT_{60}: 186, 79, or 0 ms),
- the input SNR: 10, 0, or −10 dB.

The input SNR refers to the energy ratio between a target signal and noise at the first microphone, where all sources were assumed to have the same energy.

5.2 Methods compared and evaluation metric

We compared the following ten methods differing in the way the basis vectors $\boldsymbol{e}_i(\omega)$ of $\mathscr{S}^{\perp}(\omega)$ were identified.

- Conventional MUSIC with the spatially white noise model (denoted by conv-white). $\boldsymbol{e}_i(\omega)$ were identified by EVD of $\boldsymbol{\Phi}_{\boldsymbol{xx}}$.
- Conventional MUSIC with the fixed noise coherence model in Section 3.2.2 (denoted by conv-coh). $\boldsymbol{e}_i(\omega)$ were identified by GEVD of the matrix pencil $(\boldsymbol{\Phi}_{\boldsymbol{xx}}, \boldsymbol{\Gamma})$.

[1] E. Vincent and D. R. Campbell, "Roomsimove," (2010, Nov. 29). [Online]. Available: http://www.irisa.fr/metiss/members/evincent/software.

Table 2: Precision averaged over all mixtures for $R = 2$ and $B = 16$.

method	conventional		LMC				TNM			
noise model	white	coh	uncor	coh	BND	real	uncor	coh	BND	real
precision	0.51	0.58	0.41	0.59	0.42	0.48	0.51	0.58	**0.63**	0.35

- Proposed MUSIC based on LMC with the four noise models in Section 3 (denoted by comp-uncor, comp-coh, comp-BND, or comp-real). $\boldsymbol{e}_i(\omega)$ were identified by EVD of $\boldsymbol{\Phi}_{\boldsymbol{cc}}(\tau, \omega)$ estimated by LMC.
- Proposed MUSIC based on TNM with the four noise models in Section 3 (denoted by trace-uncor, trace-coh, trace-BND, or trace-real). $\boldsymbol{e}_i(\omega)$ were identified by EVD of $\boldsymbol{\Phi}_{\boldsymbol{cc}}(\tau, \omega)$ estimated by TNM.

We estimated the target azimuths by picking the L largest peaks in the MUSIC spectrum, and assessed their *precisions* [33], which is the ratio of correct azimuth estimates and L (assumed to be known). Here, the correct estimates are defined as those within 5° from a true azimuth.

We used the cylindrically isotropic noise model [22] to calculate $\boldsymbol{\Gamma}$ in conv-coh, comp-coh, and trace-coh. We set $\omega_{\min} = 94\,\mathrm{Hz}$ and $\omega_{\max} = 8\,\mathrm{kHz}$ in (9), so as to remove the low frequency bins with extremely low SNRs. We divided the data duration into B segments, calculated the wideband MUSIC spectrum in each segment, and geometrically averaged them to obtain a single MUSIC spectrum.

5.3 Experimental results

Table 2 shows the precision for each method averaged over all mixtures. We set $R = 2$ and $B = 16$, because $R \geq 2$ resulted in a reasonable performance, and B had only a little impact in our preliminary experiments. The trace-BND algorithm gave the highest precision of 0.63 higher than that of state-of-the-art conv-coh by 0.05. The precisions of comp-coh and trace-coh were comparable to that of conv-coh based on the same noise model. The precisions of the other proposed algorithms were lower than that of conv-coh. The poor performance with the uncorrelated noise model and the real-valued noise covariance model is accounted for by high spatial correlation of real-world noise and the high model dimension, respectively. For this reason, the following evaluation is focused on the proposed trace-BND, comp-coh, and trace-coh and on the conventional conv-white and conv-coh as baselines.

Figure 2 shows examples of MUSIC spectra for two sources under a highly reverberant and noisy condition: $\mathrm{RT}_{60} = 186\,\mathrm{ms}$; SNR: $-10\,\mathrm{dB}$. The true azimuths are depicted by the vertical lines in the figure. The proposed three algorithms resulted in smaller estimation error than the conventional methods, and trace-BND resulted in the most accurate estimation.

Figs. 3 and 4 show the impact of the input SNR and reverberation. We set $R = 2$ and $B = 16$. As we see in Fig. 3, the precisions decreased when the input SNR decreased. While the precision of trace-BND was comparable to those of the methods based on the fixed noise coherence model at 10 dB, it was higher by 0.06–0.09 at 0 dB and $-10\,\mathrm{dB}$, showing that it is more robust against noise. As seen from Fig. 4, the precision of trace-BND was comparable to those of the

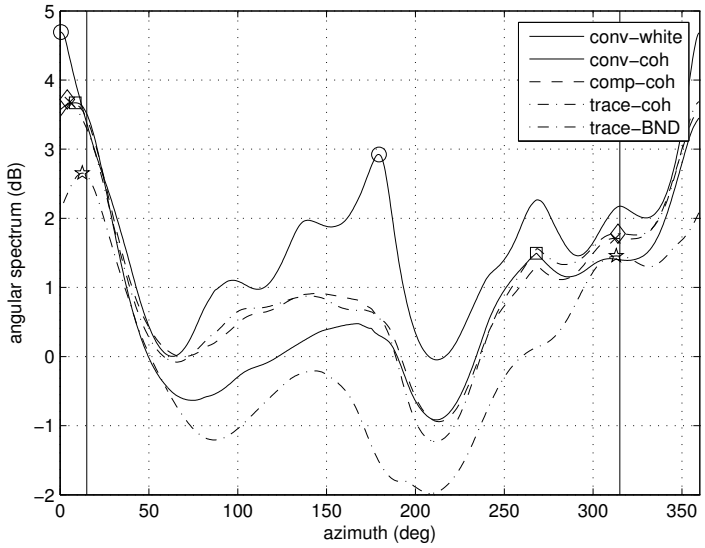


Figure 2: Examples of MUSIC spectra. $L = 2$; angle between adjacent sources: 60°; absorption coefficient: 0.4; SNR: -10 dB; $R = 2$; $B = 1$.

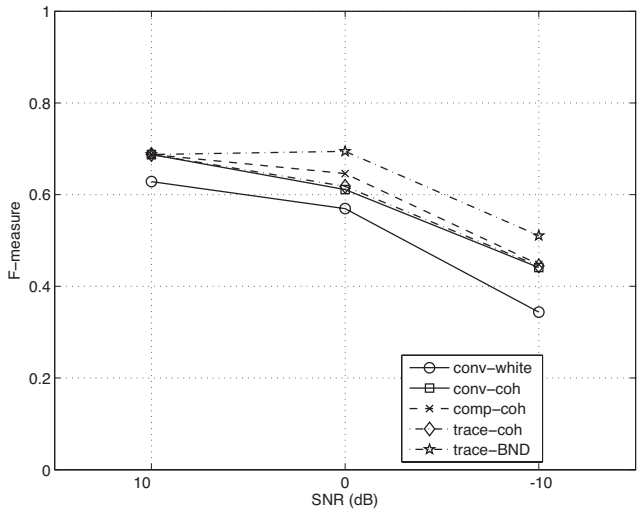


Figure 3: Average precision as a function of the SNR for the conventional and proposed methods for $R = 2$ and $B = 16$.

methods based on the fixed noise coherence model for the absorption coefficients of 1 and 0.7, whereas it was higher by 0.12–0.18 for that of 0.4. This robustness of trace-BND against reverberation is likely because late reverberation is uncorrelated to the direct path and can be regarded as diffuse noise, so that it is well explained by the BND model.

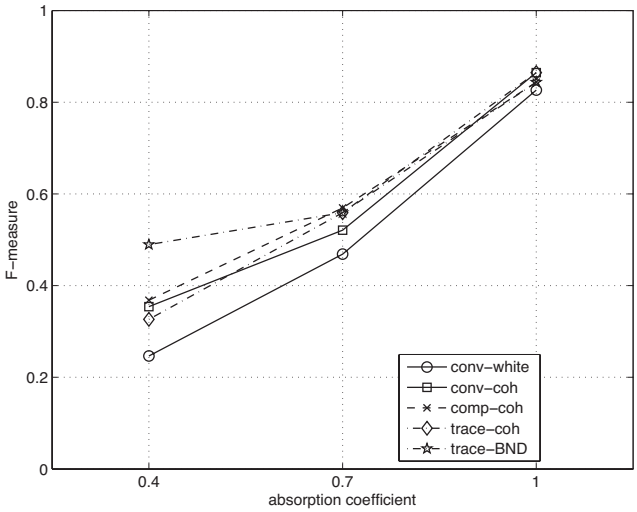


Figure 4: Precision as a function of the absorption coefficient of the walls for the conventional and proposed methods for $R = 2$ and $B = 16$.

6 Conclusion

This paper proposed a framework for robust DOA estimation of multiple target sources in diffuse noise. Our approach is based on estimating the target spatial covariance matrix from the observed spatial covariance matrix. This enables the identification of the orthogonal complement of the target signal subspace as the null space of the estimated matrix, whereby enabling accurate DOA estimation via MUSIC. We presented a unified framework for modeling noise covariance in a matrix space. This enables us to obtain a noise-free component of the observed spatial covariance matrix through projecting it onto the orthogonal complement of the noise matrix subspace. This noise model was shown to include four state-of-the-art diffuse noise models as special cases, namely the spatially uncorrelated noise model, the fixed noise coherence model, the blind noise decorrelation model, and the real-valued noise covariance model. We express the target spatial covariance matrix as the sum of two components: one belonging to the noise matrix subspace and one orthogonal to it. The latter is obtained by the projection, whereas the former is reconstructed exploiting the low-rankness of the target spatial covariance matrix. We proposed two algorithms for this matrix completion, namely the low-rank matrix completion and the trace norm minimization algorithms. We evaluated the performance of the proposed methods using a large dataset, and showed that the proposed trace norm minimization algorithm with the blind noise decorrelation model outperformed the conventional MUSIC methods essentially in terms of the precision.

Acknowledgments

This work was supported by Grant-in-Aid for JSPS Fellows 22-6927 from MEXT, Japan, and by INRIA under the Associate Team Program VERSAMUS

(http://versamus.inria.fr/)

References

[1] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, USA, Oct. 1997.

[2] M. Brandstein and D. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, Springer-Verlag, Berlin, 2001.

[3] W. Foy, "Position-location solutions by Taylor-series estimation," *IEEE Trans. Aerospace Electro. Sys.*, vol. 12, no. 2, pp. 187–194, Mar. 1976.

[4] J. Smith and J. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 12, pp. 1661–1669, Dec. 1987.

[5] P. Stoica and J. Li, "Source localization from range-difference measurements," *IEEE Signal Process. Magazine*, vol. 23, no. 6, pp. 63–69, Nov. 2006.

[6] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[7] Y. Izumi, N. Ono, and S. Sagayama, "Sparseness-based 2ch BSS using the EM algorithm in reverberant environment," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2007, pp. 147–150.

[8] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Grouping separated frequency components by estimating propagation model parameters in frequency-domain blind source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 5, pp. 1592–1604, 2007.

[9] M. Mandel, D. Ellis, and T. Jebara, "An EM algorithm for localizing multiple sound sources in reverberant environments," in *Proc. Neural Information Processing Systems (NIPS)*, 2006, pp. 953–960.

[10] J. Dibiase, H. Silverman, and M. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein and D. Ward, Eds., chapter 8. Springer, 2001.

[11] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[12] K. Nakadai, H. Nakajima, M. Murase, H.G. Okuno, Y. Hasegawa, and H. Tsujino, "Real-time tracking of multiple sound sources by integration of in-room and robot-embedded microphone arrays," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct. 2006, pp. 852–859.

[13] N. Ito, E. Vincent, N. Ono, R. Gribonval, and S. Sagayama, "Crystal-MUSIC: Accurate localization of multiple sources in diffuse noise environments using crystal-shaped microphone arrays," in *Proc. of LVA/ICA, Lecture Notes in Computer Science*, Saint-Malo, France, Sept. 2010, vol. 6365, pp. 81–88.

[14] A. Paulraj, R. Roy, and T. Kailath, "A subspace rotation approach to signal parameter estimation," *Proc. IEEE*, vol. 74, no. 7, pp. 1044–1046, 1986.

[15] M. Wax, T. Shan, and T. Kailath, "Spatio-temporal spectral analysis by eignestructure methods," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 4, pp. 817–827, Aug. 1984.

[16] Richard K. Cook, R. V. Waterhouse, R. D. Berendt, Seymour Edelman, and M. C. Thompson Jr., "Measurement of correlation coefficients in reverberant sound fields," *J. Acoust. Soc. Am.*, vol. 27, no. 6, pp. 1072–1077, Nov. 1955.

[17] Rainer Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Proc. IEEE Int'l Conf. Acoust. Speech Signal Process. (ICASSP)*, New York, Apr. 1988, pp. 2578–2581.

[18] Iain A. McCowan and Hervé Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 709–716, Nov. 2003.

[19] N. Ito, H. Shimizu, N. Ono, and S. Sagayama, "Diffuse noise suppression using crystal-shaped microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2101–2110, Sept. 2011.

[20] Hikaru Shimizu, Nobutaka Ono, Kyosuke Matsumoto, and Shigeki Sagayama, "Isotropic noise suppression in the power spectrum domain by symmetric microphone arrays," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, Oct. 2007, pp. 54–57.

[21] Nobutaka Ito, Nobutaka Ono, and Shigeki Sagayama, "Designing the Wiener post-filter for diffuse noise suppression using imaginary parts of inter-channel cross-spectra," in *Proc. IEEE Int'l Conf. Acoust. Speech Signal Process. (ICASSP)*, Mar. 2010, pp. 2818 – 2821.

[22] Gary W. Elko, "Spatial coherence functions for differential microphones in isotropic noise fields," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein and D. Ward, Eds., chapter 4, pp. 61–85. Springer-Verlag, Berlin, 2001.

[23] N. Srebro and T. Jaakkola, "Weighted low-rank approximations," in *Proc. International Conference on Machine Learning (ICML)*. AAAI Press, 2003, pp. 720–727.

[24] K. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of Optimization*, vol. 6, no. 3, pp. 615–640, Sept. 2010.

[25] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *The Journal of the Society for the Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, Apr. 2009.

[26] S. Ji and J. Ye, "An accelerated gradient method for trace norm minimization," in *Proc. International Conference on Machine Learning (ICML)*, Montreal, Canada, 2009, pp. 457–464.

[27] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, Sept. 1936.

[28] D. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

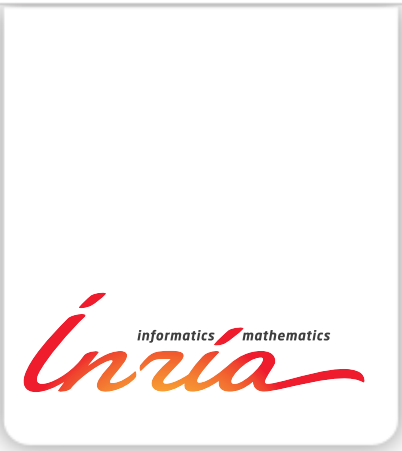
[29] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 56, pp. 381–389, 2009.

[30] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Mathematical Programming*, vol. 128, pp. 321–353, 2011.

[31] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[32] Akira Kurematsu, Kazuya Takeda, Yoshinori Sagisaka, Shigeru Katagiri, Hisao Kuwabara, and Kiyohiro Shikano, "ATR Japanese speech database as a tool of speech recognition and synthesis," *Speech Commun.*, vol. 9, no. 4, pp. 357–363, Aug. 1990.

[33] Charles Blandin, Alexey Ozerov, and Emmanuel Vincent, "Multi-source TDOA estimation in reverberant audio using angular spectra and clustering," *Signal Processing*, vol. 92, pp. 1950–1960, 2012.



RESEARCH CENTRE
RENNES – BRETAGNE ATLANTIQUE

Campus universitaire de Beaulieu
35042 Rennes Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399